1 Supplement: Benchmarking Profiles for Representative SBML Models

We have benchmarked the JIT-based libRoadRunner and the popular interpreter-based Jarnac network solvers on a wide variety of typical biological network models:

- *AMPA Receptor*: A neuronal model which calculates ionic current on the basis of 16 kinetic states. This model contains 19 SBML species and 21 reactions.
- *Jana-Wolf Glycolysis*: A glycolytic model featuring complex dynamics with 15 species, 11 reactions, and $\sim$ 30 parameters with complex rate laws
- *Biomodels Model BIOMD0000000014*: A large, mass-action model of a 3-stage mitogen-activated protein kinase (MAPK). Contains 86 species and 300 reactions.
- *Biomodels Model: BIOMD0000000033*: A model which makes heavy use of Michaelis-Menten kinetics. Contains 32 species and 26 reactions.
- A series of Brusselator models, each containing N copies of a single Brusselator. We increase the number of copies in increments of 50 up to a maximum of 500, with the smallest model having 8 species and 8 reactions and the largest having 1006 species and 2004 reactions. Since the copies are independent, this benchmark tests the simulator's ability to compute rates of change in a large sparse matrix
- A simple stochastic model.

2 Benchmarking Method

We ran all benchmarks on a workstation running Windows 8.1 Enterprise 64-bit with an Intel Core 2 Quad Q9550 at 2.83 GHz and 4 GB of RAM and a Samsung HD103UJ ATA drive for storage. We replicated all benchmarks three times on the same computer and averaged the results. We used libRoadRunner version 1.3.1 and Jarnac version 3.33b.

2.1 ODE Solver

libRoadRunner used bench.py, a Python script which iterates over each model in the benchmarking suite, executing and timing each model. Benchmarks for Jarnac used jarnac_bench.jan, a Jarnac script which iterates over each model in the benchmarking suite, executing and timing each model.

Model	libRoadRunner (s)	Jarnac (s)	Jarnac/libRoadRunner
AMPA Receptor	0.000	0.094	N/A
JanaWolf	0.187	0.994	5.38
BIOMOD14	0.007	2.28	324
BIOMOD33	0.042	0.176	4.07
Brusselator1	0.001	0.015	15.0
Brusselator2	0.001	0.021	16.0
Brusselator50	0.004	2.47	617
Brusselator100	0.007	12.7	1819
Brusselator150	0.011	35.5	3553
Brusselator200	0.014	76.7	5112
Brusselator250	0.023	142	7905
Brusselator300	0.022	245	11141
Brusselator350	0.026	375	14560
Brusselator400	0.033	544	16936
Brusselator450	0.035	759	18211
Brusselator500	0.045	1026	26413
Stochastic	0.000	7.78	N/A

Table S.1: Comparison of libRoadRunner and Jarnac run times for a suite of benchmark models.

2.2 Stochastic (Gillespie) Solver

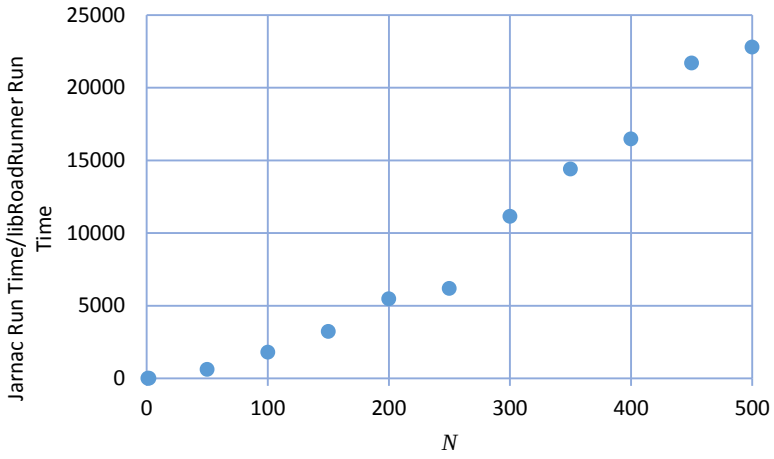
libRoadRunner and Jarnac include stochastic solvers based on Gillespie's Direct Method. The scripts bench_stoch.py and jarnac_bench_stoch.jan benchmark the stochastic model on libRoadRunner and Jarnac respectively.

3 Results

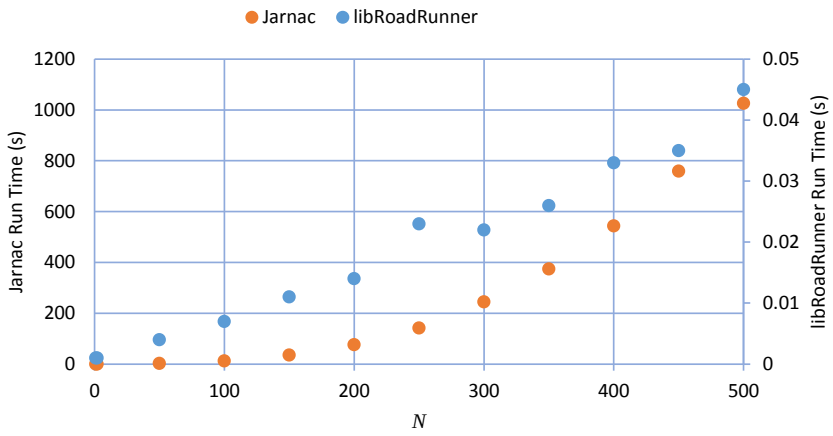
Table S.1 shows the numeric results of the benchmark. Figure S.1 shows the Brusselator results. libRoadRunner run times scale linearly with the number of replicas ($t \sim N$). Jarnac run times scale as $t \sim N^{2.6}$. The linear scaling in model size makes libRoadRunner suitable for solution of large multi-scale simulations which include multiple replicas of complex SBML models.

4 Comments

The Jarnac ODE benchmark results in Table S.1 come from an initial run including all models except the AMPA receptor and the Brusselator450/500 models. Jarnac reported the error `<ESymNotInitialised>: Error: Symbol [conduc_02] has not been initialised with a value` when loading the SBML for the AMPA receptor model. Using Jarnac's built-in converter to convert the



(a)



(b)

Figure S.1: (a) Ratio of Jarnac to libRoadRunner run times (t) vs. number of model replicas (N) for the Brusselator model.(b) Jarnac (Orange dots, left hand scale) and libRoadRunner (Blue dots, right hand scale) run times (t) vs. number of model replicas (N) for the Brusselator model.

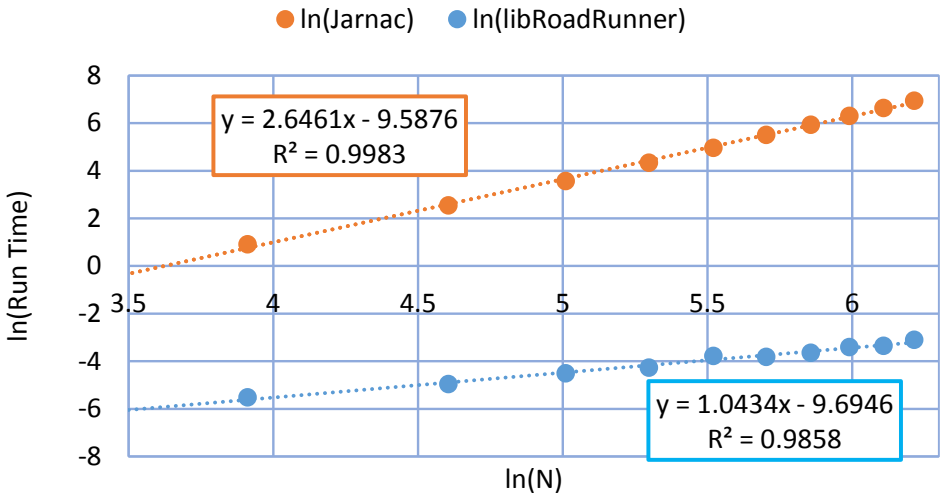


Figure S.2: Scaling of Jarnac (Orange dots) and libRoadRunner (Blue dots) run times (t) vs. number of model replicas (N) for the Brusselator model. Linear fits (Dashed lines) of ln(Run Time) vs. $ln(N)$ show that the run times scales as $t \sim N^{2.6}$ in Jarnac and $t \sim N$ in libRoadRunner.

model from SBML into a Jarnac script eliminated this error. Jarnac exhausted its available memory on the third trial while executing Brusselator models 450 and 500. We therefore re-ran these benchmarks independently (for Jarnac only) and included their timings in Table S.1. These independent runs produced timings comparable to those in the test suite. libRoadRunner completed some performance tests (such as the AMPA receptor) faster than the measured resolution. The *Jarnac/libRoadRunner* run time ratio is reported as "N/A" for these cases.

5 Availability

Python and Jarnac scripts used to run the benchmarks are included as additional supplementary data. The full source code for the benchmarking suite is available at https://github.com/0u812/rr-benchmarking.